Practical Issues with Intrusion Detection

Intrusion Detection — How?

- Where do sensors go?
- How do you put them there?
- Sensor issues
- Other techniques
- Ethical and legal issues

Sensors

Locations

- Outside the firewall?
- We *know* there are bad guys there; what's the point?
- Just inside? What's the threat model?
- On sensitive internal nets?
- In front of each sensitive host?
- In "dark space"?

What's Dark Space?

- A block of address space not used by real machines and not pointed to by DNS entries
- There is no legitimate reason to send packets to such addresses
- Therefore, any host sending to such addresses is up to no good
- Commonly used to detect scanning worms

What's the Purpose?

- Unless you're a researcher, you care about real threats to your own machines
- Inside the firewall? Detect data exfiltration
- Sensitive internal nets: detect threats aimed at them
- Watching each host? Detect attacks on inside hosts from other hosts on the same LAN
- Dark space? Detect scanning worms (and attackers)

Auto-Quarantine

- Many organizations implement "auto-quarantine"
- This is especially common for university residence hall networks
- Machines that do too much scanning (and in particular attempt to probe dark space) are assumed to be virus-infected
- They're moved to a separate net; the only sites they can contact are Windows Update, anti-virus companies, and the like

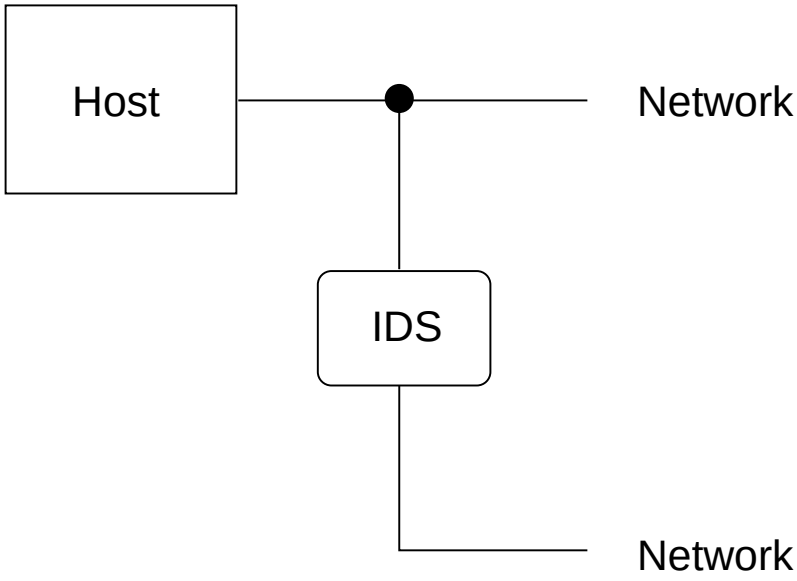
Honeypots and Honeynets

- Special-purpose host or network designed to be attacked
- Equipped with copious monitoring
- Lure the attacker in deeper
- Waste the attacker's time; study the attacker's technique
- Note well: keeping honeypot (and dark space) addresses secret is vital

Host- or Net-Resident?

- Suppose you want to monitor each host. Where does the monitor live?
- Dedicated in-line hardware: good, but expensive
- On the host: cheap, but subvertible

Net-Resident: Parallel

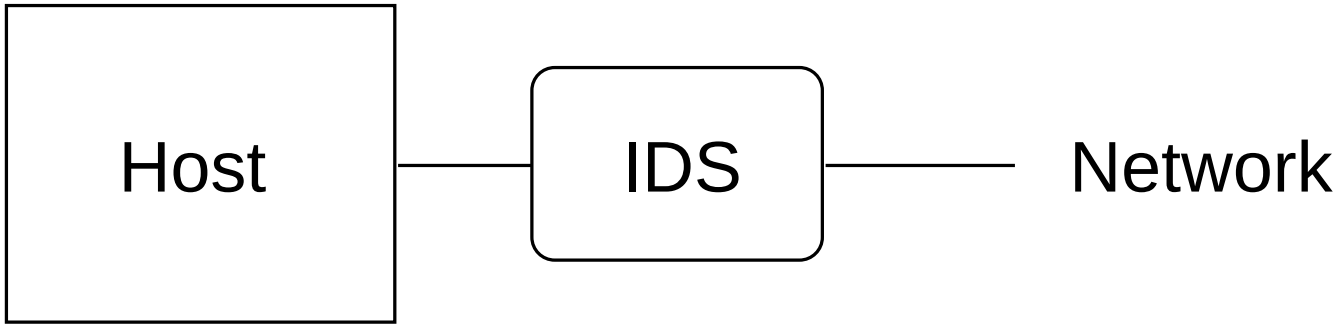


- Very unobtrusive
- But — need special hardware to tap an Ethernet
- Need some network connection to the IDS

Tapping an Ethernet

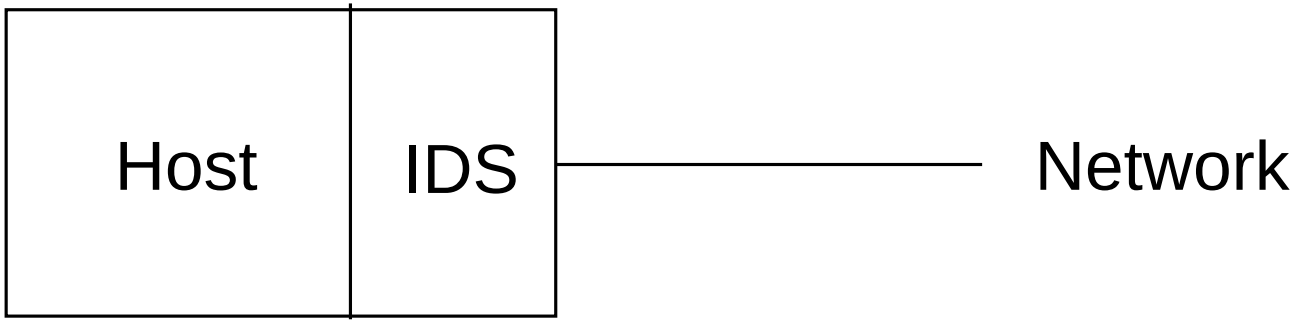
- Cannot simply wire IDS to jack
- Best solution: one-way tap gear
- Note: unidirectional only; may need a pair of them
- Some switches have a monitoring port (AKA spanning port, mirroring port, etc) — can receive copies of data from any other port
- For 10BaseT nets, use a *hub* instead of a switch

Net-Resident: Serial



- Can't miss packets
- But — if it crashes, the host is unreachable
- More detectable, via timing
- Can the IDS box be hacked?

Host-Resident Monitor



- No special hardware needed
- IDS sees exactly what host sees
- But — subvertible
- Useful precaution: immediately transmit IDS data elsewhere

TCP Normalization

- Attackers can play games with TCP/IP to confuse network-resident IDS
- Example: overlapping fragments:

```
s   u       n   o   r   m
            r   o   o   t
```

  Which fragment is honored?
- TTL games: give some packets a TTL just high enough to reach the IDS, but not high enough to reach the destination host
- Solution: *TCP normalizer*, to fix these

The Big Advantages of Host IDS

- More time
- More context
- Everything is reassembled
- Look at entire item, not streams
- Example: it's all but impossible to do email virus scanning in the network

Extrusion Detection

- Detect bad things leaving your network
- Detect sensitive things leaving your network
- Finds theft of inside information, either by attacker or by rogue insider
- Can be done in the network or in application gateways

Simple Logging

Simple Logging

- I ran this command for a while, on two hosts:

```
tcpdump -p -l "tcp[13] == 0x2 and dst $us"
```

- What does it do?
- Logs all TCP SYN-only packets addressed to us (tcp[13] is the flags byte in the TCP header; 0x2 is SYN)

Some Results

- About 85 probes apiece, during a 30-hour run
- 63 different ports scanned
- Some obvious: http, ssh, Windows file-sharing, SMTP, web proxy
- Some strange: 49400–49402, 8081–8090, 81–86
- Some ominous: terabase, radmin-port
- Most probers looked at one port; one looked at 46 ports

The Most Probed Ports

Scans	*Port*
3	ms-wbt-server
3	ssh
5	8000
5	http-alt
6	ms-sql-s
6	radmin-port
7	BackupExec
8	smtp
9	WebProxy
9	http

What Did The Probers Want?

- WebProxy and SMTP are probably for spam email and connection-laundering
- The others look like probes for known vulnerabilities
- http could have been a "spider" or it could be looking for known holes

Broader Data

- Useful source: `http://www.dshield.org`
- Its current Top 10 list shown at right
- Clearly, the probers are interested in peer-to-peer servers...
- Some ports are mysterious

Name	*Port*
—	15281
win-rpc	1026
eDonkey2000	4662
eMule	4672
icq	1027
bittorrent	6881
—	1028
gnutella-svc	6346
smtp	25
microsoft-ds	445

Bad Neighborhoods

- I see more probes here than elsewhere. Why?
- There are different "neighborhoods" — ranges of IP addresses — in cyberspace
- University networks are good hunting — few firewalls, good bandwidth, many poorly-administered machines
- Newly-allocated network blocks have few hosts, and aren't scanned as much

Log Files

Shadow Hawk

Shadow Hawk Busted Again

As many of you know, Shadow Hawk (a/k/a Shadow Hawk 1) had his home searched by agents of the FBI...

When he was tagged by the feds, he had been downloading software (in the form of C sources) from various AT&T systems. According to reports, these included the Bell Labs installations at Naperville, Illinois and Murray Hill, New Jersey.

—Phrack Issue 16, File 11, November 1987

How was Shadow Hawk Detected?

- He had broken into some Bell Labs machines
- He tried to use `uucp` — a dial-up file transfer/email system that came with Unix — to grab `/etc/passwd` files from other machines
- Uucp logged all file transfer requests
- Several people at Murray Hill had automated jobs that scanned the log files for anything suspicious

Stalking the Wily Hacker

- An accounting file didn't balance — a username had been added without the proper bookkeeping entries
- Cliff Stoll noticed and tried to figure out what was going on
- Ultimately, it led to a KGB-controlled operation aimed at military secrets...

What was the Common Thread?

- Log files of various sorts
- "Extraneous" information
- Log files can prevent problems, help you figure out how the system was penetrated, what was affected, and — if you're lucky and persistent — who did it

Where Do Log Files Come From?

- Many different system components can produce logs
- Often, these aren't enabled by default
- Should they be?

Detecting Problems Via Logfiles

The "Code Red" worm activity can be identified on a machine by the presence of the following string in a web server log files:

```
/default.ida?NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNN%u9090%u6858%ucbd3%u7801%
u7801%u9090%u6858%ucbd3%u7801%u9090%u9090%u8190%u00c
b%u53ff%u0078%u0000%u00=a
```

From

http://www.cert.org/advisories/CA-2001-19.html

An Attempted Intrusion?


```
[Sun Nov 20 23:17:18 2005] [error] [client www.xxx.y
File does not exist: /usr/pkg/share/httpd/htdocs/xml
[Sun Nov 20 23:17:28 2005] [error] [client www.xxx.y
File does not exist: /usr/pkg/share/httpd/htdocs/php
```

(There were many more attempts from that IP address.) Both of these represent services with known security holes

Problems with Log Files

- How did I spot those probes?
- Manual search through error_log
- Not very scalable...

CS@CU

Log File Scanners

- Need to automate scans
- Pick out "interesting" events
- Hmm — what's interesting?

Log Files and Intrusion Detection

- Analyzing log files like that is a form of intrusion detection
- Can look for specific signatures, such as examples above
- Or — can look for anomalous patterns, such as too many misses or too-long URLs

Correlating Log Files

- Sometimes, the interesting information is spread among several log files
- Need accurate timestamps for correlation between machines
- Timestamps should generally be in UTC, rather than the local timezone

Types of Correlation

- Intra-machine — different forms of logfile
- Intra-site
- Inter-site
- Watch out for privacy issues!

Finding Compromised Hosts

Finding Compromised Hosts

- Suppose you've identified a compromised host. Now what?
- Get data: IP address and (when feasible) MAC address
- Find it

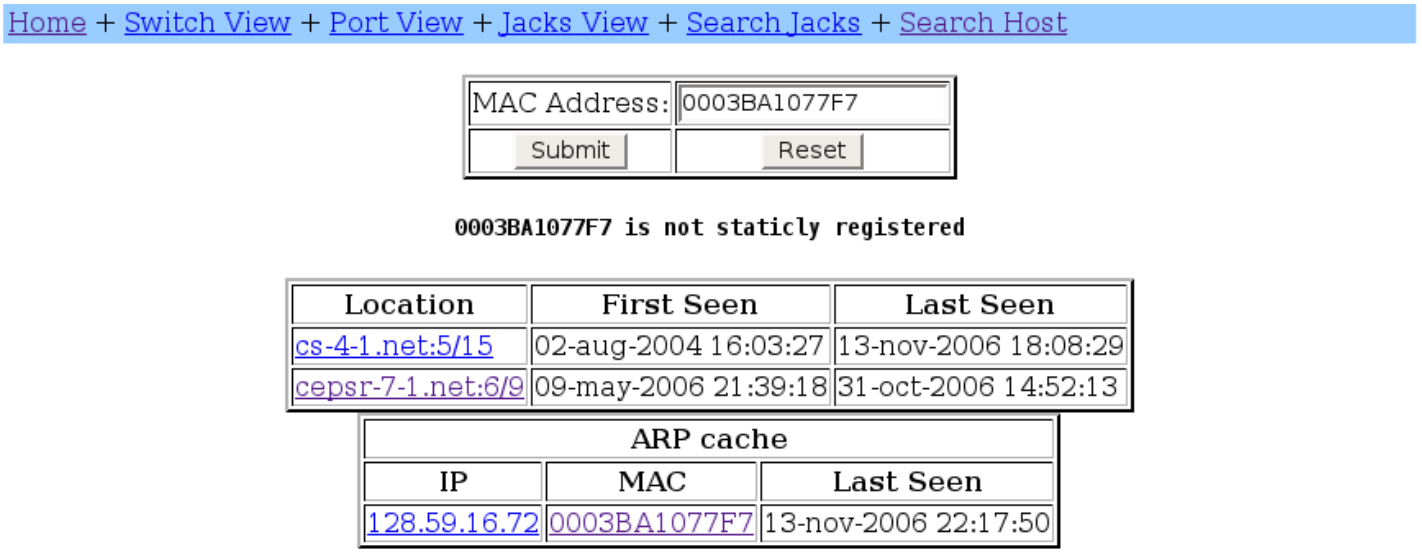
Databases

- Must be able to map IP address to location
- Must be able to map IP address to person
- Difficult on this campus — wide-open nets
- Primary reason for host registration in many places

Layer 2 Data

- Enterprise-grade switches are "managed"
- They can map an IP address or a MAC address to a physical port
- Especially useful if the attacker is forging addresses...

Switch Data



Home + Switch View + Port View + Jacks View + Search Jacks + Search Host

MAC Address: 0003BA1077F7

Submit Reset

0003BA1077F7 is not staticly registered

Location	First Seen	Last Seen
cs-4-1.net:5/15	02-aug-2004 16:03:27	13-nov-2006 18:08:29
cepsr-7-1.net:6/9	09-may-2006 21:39:18	31-oct-2006 14:52:13

ARP cache

IP	MAC	Last Seen
128.59.16.72	0003BA1077F7	13-nov-2006 22:17:50

Note that a single MAC address has shown up on two different switch ports, in different buildings. This is reasonable for a laptop, but not for a server!

Locating an Evil WiFi Laptop

- Ask the switch what access point it's near
- Ping-flood the machine
- Wander around the room looking at the lights...